Theodore Johnson
University of Florida

ted@cis.ufl.edu

Ben Kobler NASA GSFC
Chris Daly NASA GSFC
Bob Howard Hughes AIS

Jeanne Behnke NASA GSFC
Joe King NASA GSFC

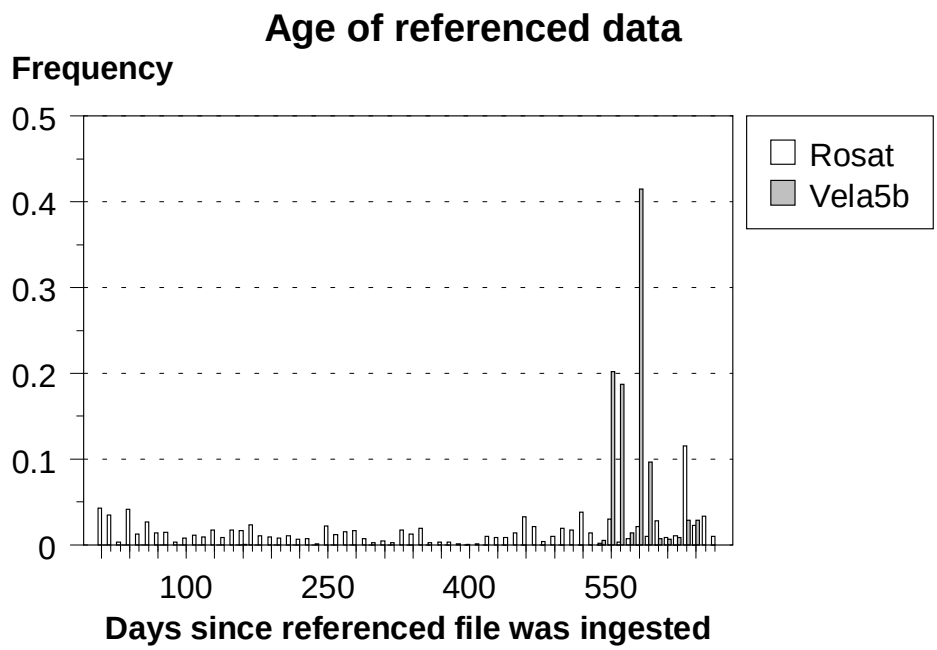
Jean Bedet Hughes STX

Bill North NASA GSFC
Joan Bixby Dunham Computer Sciences Corp.

Carlos Guerra University of Florida

- New data has the highest rate
  But, most accesses are to old data.
- Requests occur in batches
- Caching can be effective
- High percentage of repeat requests due to original requester.

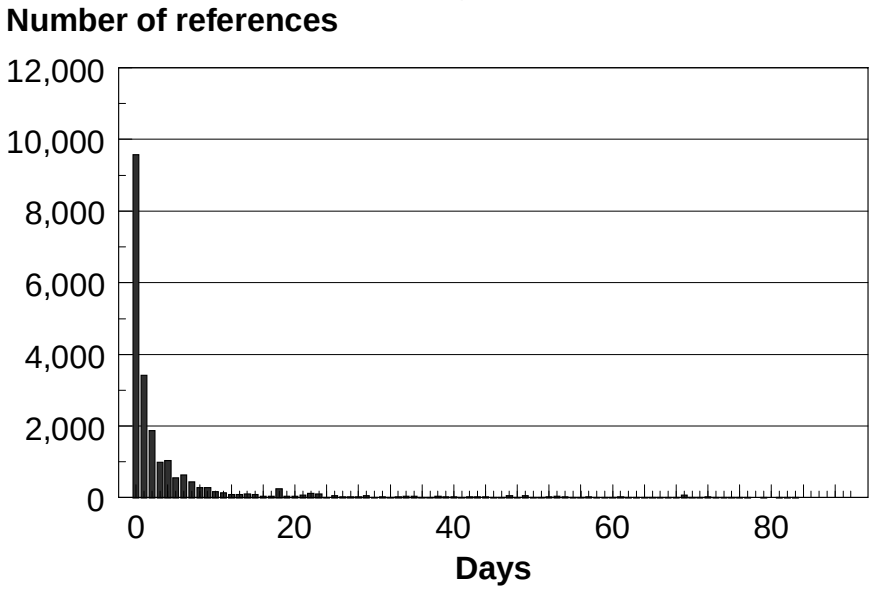
- In-depth study of GSFC V0 DAAC logs, with Jean-Jacques Bedet of Hughes STX



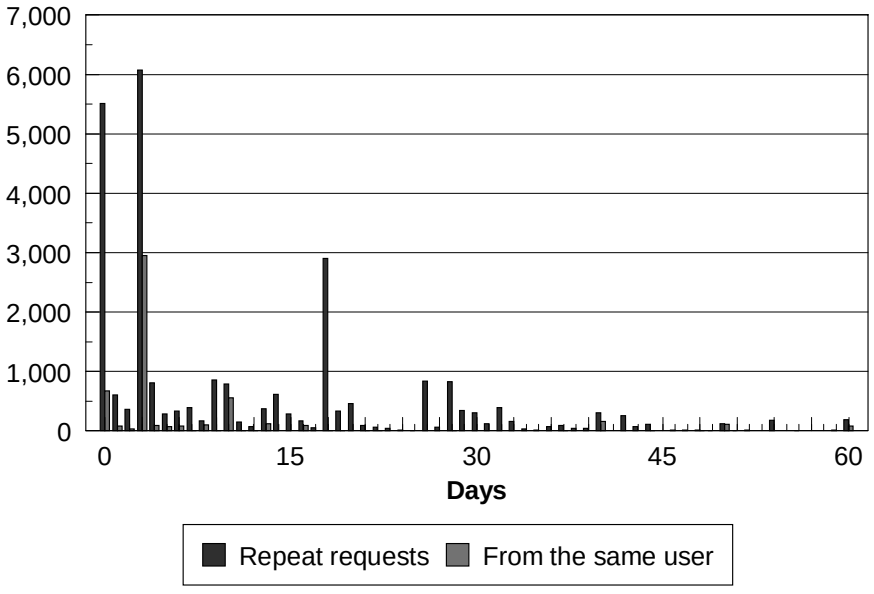
From NSSDC log files 1/95 - 4/95

Inter-reference time, 7/95 - 9/95

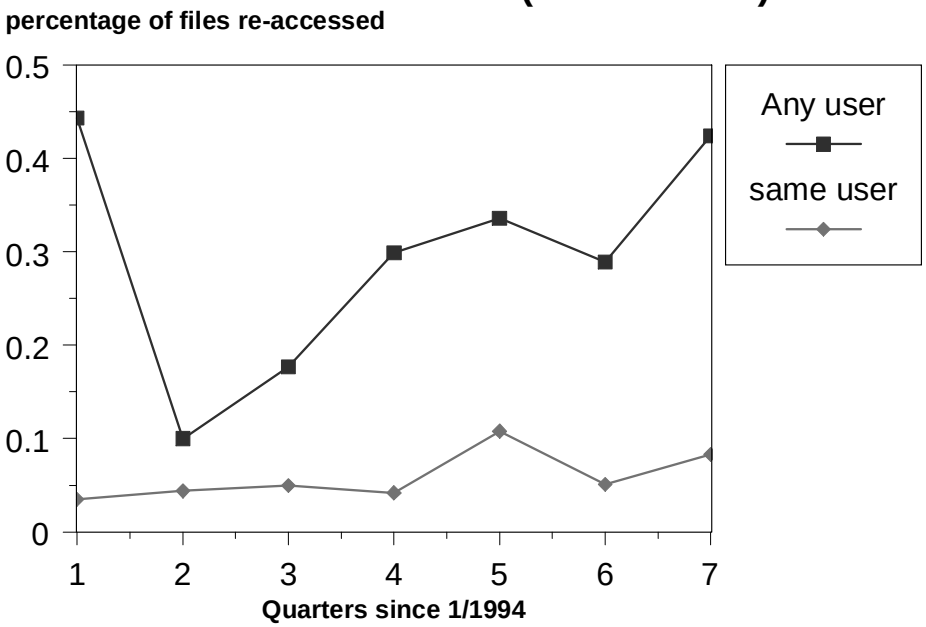
FTP access, all DAACS



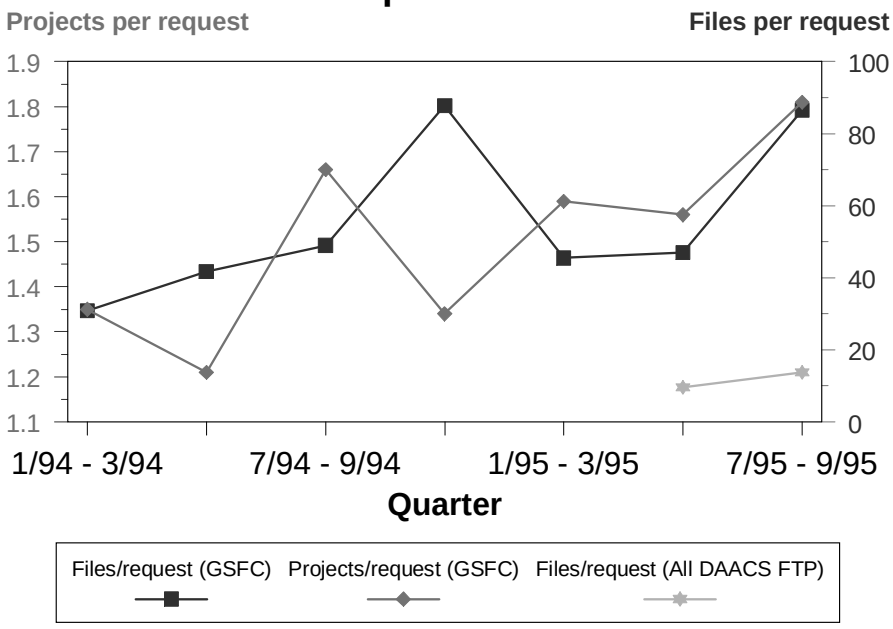
Inter-reference time, 7/95 - 9/95 (GSFC DAAC)



File re-access rates (GSFC DAAC)

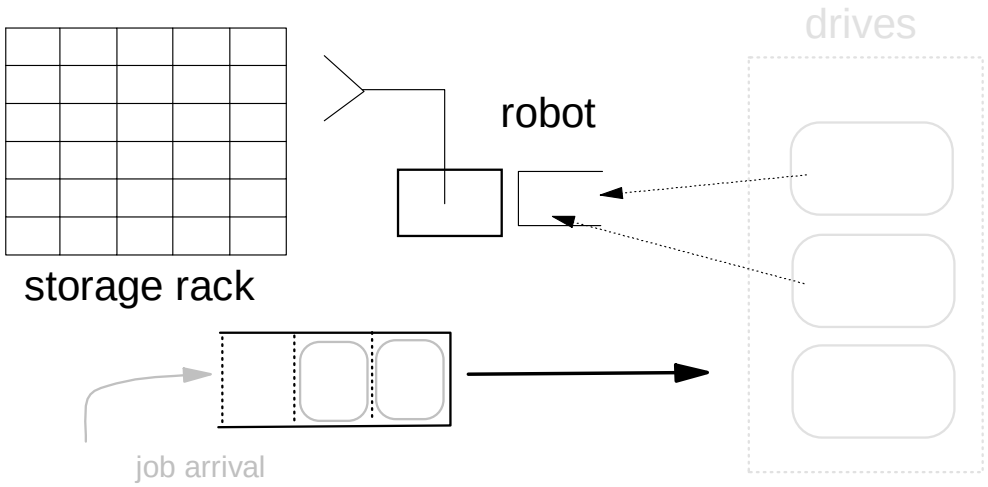


Request sizes



Mass Storage Jukebox Performance Model Available at

http://www.cis.ufl.edu/~ted/

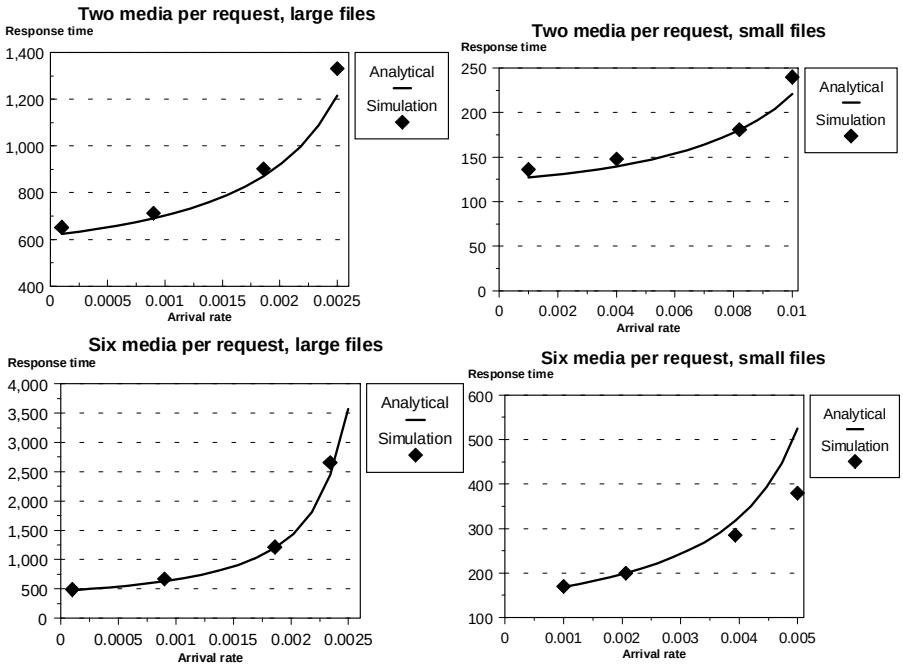


Model implementation

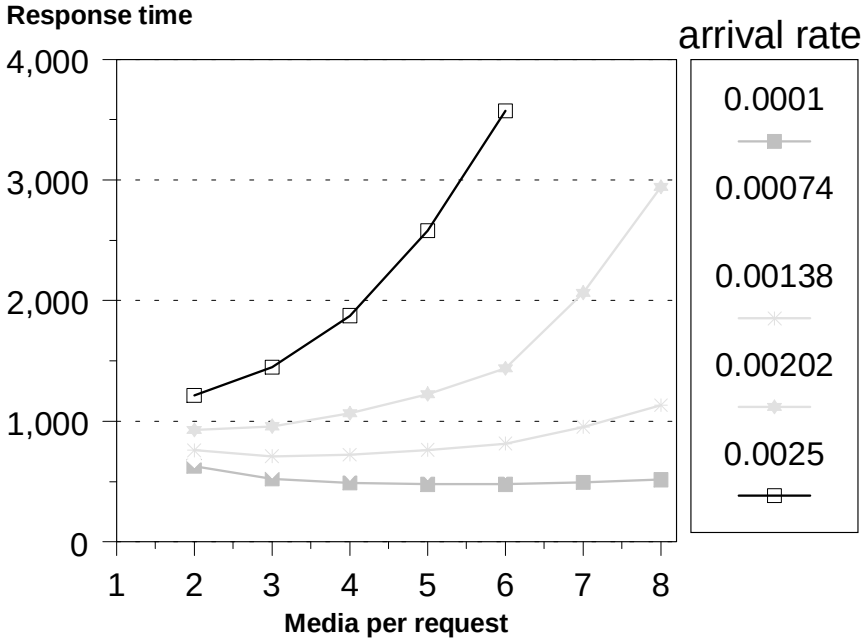
- The mass storage jukebox receives batch jobs, and has multiple drives.
- Model it as an $M^X/G/c$ queue -- solutions in the literature.
- The robot arm also sees batch arrivals, but a different distribution than for the drives.
- Compute time until the last file is transferred.
- Implemented as a stand-alone C routine.

- Geometric distribution on files, media per request
- Simple model of seek times
- Returns batch response time and drive utilization.

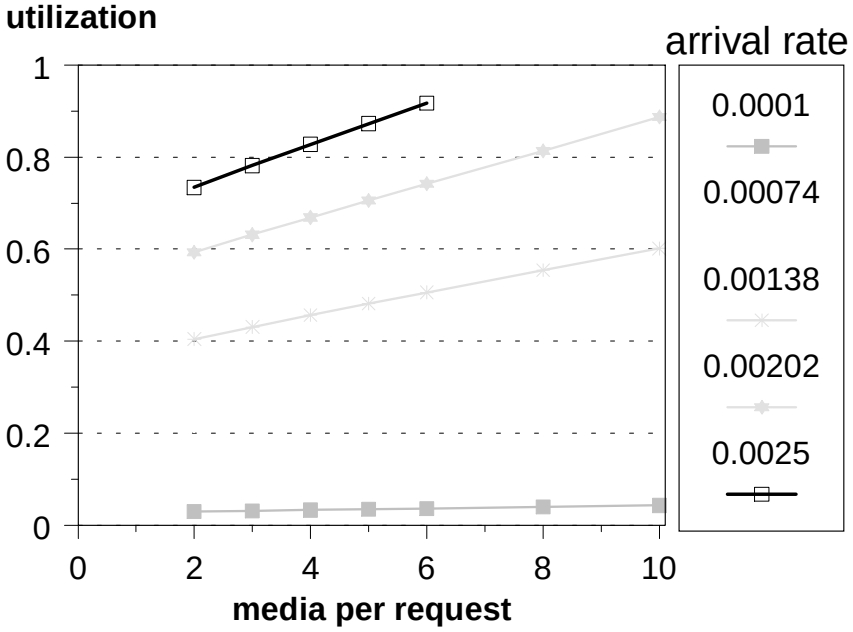
- Much faster than version 1
- Allows user-specified finite distribution of media per request
- User-specified seek time model
- Computes waiting time and variance
- Incorporated into an approximate mean value analysis queuing network model.



Varying arrival rate, large files

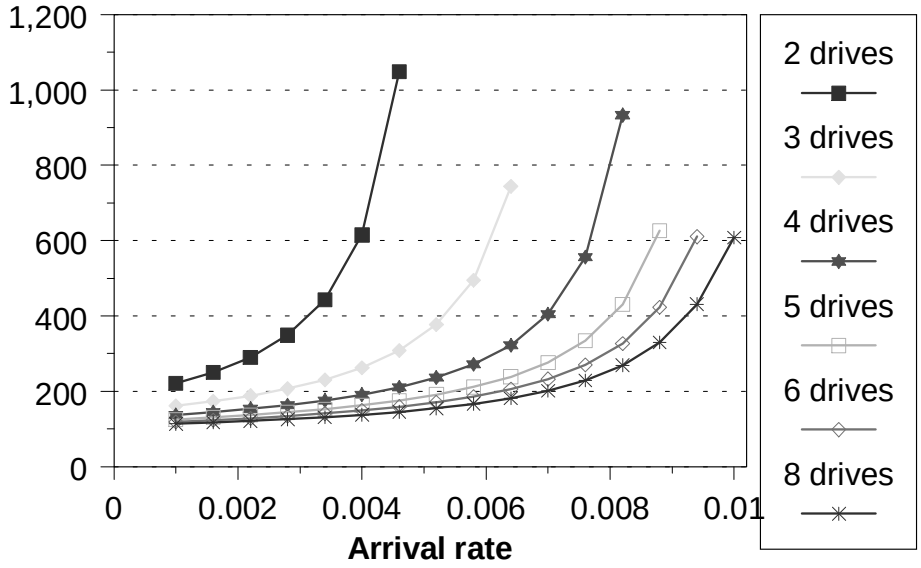


Varying arrival rate, large files



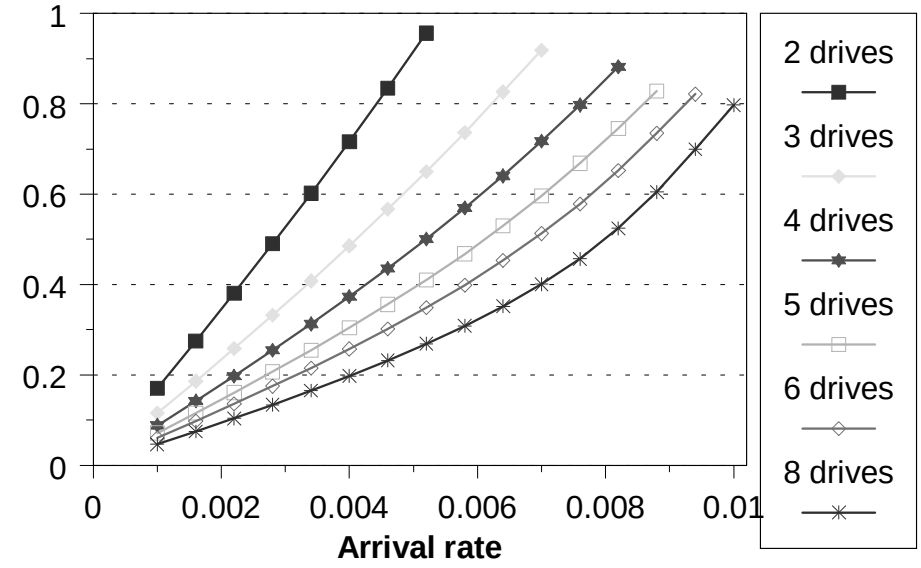
Four media per request, small files

Response time



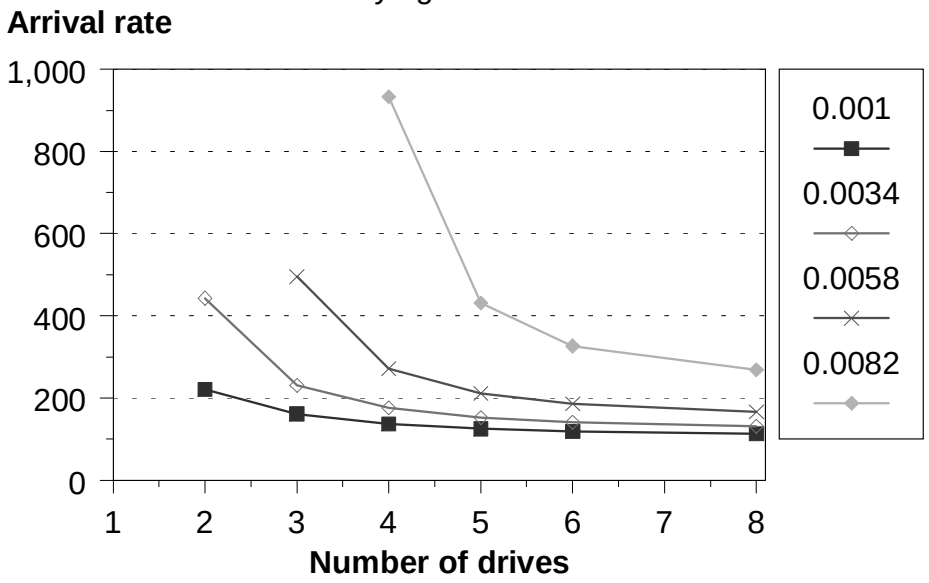
Four media per request, small files

Utilization



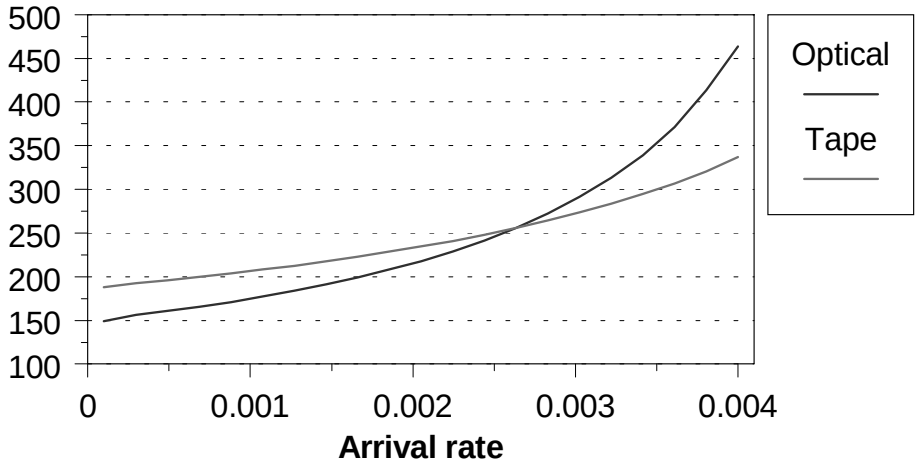
Response time vs. number of drives,

varying arrival rate.



Performance study: Optical vs. Tape drives

(hypothetical parameters)



Optical: 2 drives, 6 media per request, 1 Mbyte/sec transfer rate

Tape: 4 drives, 4 media per request, 2.5 Mbyte/sec transfer rate

Conclusions

- Continuing log file studies
- Validated mass storage jukebox performance model
- Software available at http://cis.ufl.edu/~ted/

- Improve speed, accuracy
- Increase flexibility of distributions
- Incorporate into system performance models.